SOME PROPERTIES OF MARKOFF CHAINS

BY

DAVID HAROLD BLACKWELL

A.B., University of Illinois, 1938
A.M., University of Illinois, 1939

AN ABSTRACT OF A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN MATHEMATICS IN THE GRADUATE SCHOOL OF THE UNIVERSITY OF ILLINOIS, 1941

URBANA, ILLINOIS
1941

SOME PROPERTIES OF MARKOFF CHAINS

BY

DAVID HAROLD BLACKWELL

A.B., University of Illinois, 1938
A.M., University of Illinois, 1939

AN ABSTRACT OF A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN MATHEMATICS IN THE GRADUATE SCHOOL OF THE UNIVERSITY OF ILLINOIS, 1941

URBANA, ILLINOIS
1941

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41549820_0034

SOME PROPERTIES OF MARKOFF CHAINS

Introduction. The purpose of this paper is to investigate some special types of Markoff processes. In section I we define a Markoff process and give as examples three types of Markoff processes. Later sections are devoted to a study of special cases of these three types.

I. Preliminary definitions and theorems. Let X be any abstract space and let B be a Borel field of subsets of X. Let Ω be the space of all sequences $\omega:(\dots,x_{-n},\dots,x_0,\dots,x_n,\dots)$ where $x_i \in X$ and let B' be the Borel field of subsets of Ω determined by all subsets of the form $\{x_i \in E\}$ where $E \in B$. If P(S) is any non-negative completely additive set function defined on B(B') such that $P(X)=1$ ($P(\Omega)=1$), P(S) is called a probability measure, and the space Ω with a probability measure defined on B' is called a stochastic process.

We shall later use the 1-1 transformation T of Ω into itself defined by the equation $x_n(T\omega)=x_{n+1}(\omega)$ where $x_n(\omega)$ denotes the nth coordinate of the point ω. If T is measure preserving, Ω is said to be temporally homogeneous.

If S is a set whose characteristic function depends only on $x_{a_1},x_{a_2},\dots$ where $a_1,a_2,\dots$ is a finite or infinite set of integers S is said to be a cylinder set over $x_{a_1},x_{a_2},\dots$.

If $f(\omega)$ is any integrable function defined on the process Ω and A is any Borel field included in B', there exists, by the Radon-Nikodym theorem a function $g(\omega)$, measurable with respect to A and determined uniquely up to a set of measure zero, such that for every $S \in A$, $\int_S f(\omega)dP=\int_S g(\omega)dP$. The function $g(\omega)$ is called the conditional expectation of f with respect to A. If A is the Borel field of all sets of B' which are cylinder sets over $x_{a_1},x_{a_2},\dots$, $g(\omega)$ is called the conditional expectation of f with respect to $x_{a_1},x_{a_2},\dots$ and is written $E(x_{a_1},x_{a_2},\dots;f)$. If f is the characteristic function of a set R, $E(x_{a_1},x_{a_2},\dots;f)$ is also called the conditional probability of R, with respect to $x_{a_1},x_{a_2},\dots$ and is written $P(x_{a_1},x_{a_2},\dots;R)$.

If Ω is a stochastic process whose conditional probability functions can be defined so as to satisfy $P(x_m,\dots,x_n;S)=P(x_n;S)$ for every cylinder set S over $x_{m+1},x_{m+2},\dots$, Ω is called a Markoff process.

Suppose that for each integer n and each $x \in X$ we are given a probability measure $P_n(x;E)$ defined on B. If for each n there exists a probability measure $Q_n(E)$ defined on B and satisfying the equation $Q_{n+1}(E)=\int P_n(x;E)dQ_n$, the functions $Q_n(E)$ are called a set of absolute probabilities corresponding to the functions $P_n(x;E)$.

Theorem 1: Given any set of functions $P_n(x;E)$ as described above for which there exists a corresponding set of absolute probabilities $Q_n(E)$, there exists a uniquely determined stochastic process Ω such that

(a) $P(x_m,\dots,x_n;\{x_{m+1} \in E\})=P_n(x_n;E)$ almost everywhere

(b) $P\{x_n \in E\}=Q_n(E)$.

This process Ω is a Markoff process.

Theorem 2: Let B(B') be a Borel field of sets in a space X(Y) and suppose that for each $y \in Y$, m(y;E) is a measurable function of y. If M(E) is any measure defined on B' and if $m(E)=\int m(y;E)dM(y)$, then for every function f(x) integrable with respect to m, $\int f(x)dm(x)=\int(\int f(x)dm(y;x))dM(y)$.

Corollary 1: For all x we have the equation $\int P_{n+m}^{(s)}(x;E)dP_n^{(s)}(x_0;x)=P_m^{(s+n)}(x_0;E)$" where $P_n^{(s)}(x;E)$ is defined in-

---

' IX, p. 36.

" $\int dP_n^{(s)}(x_0;x)$ denotes integration of the variable x with respect to $P_n^{(s)}(x_0;E)$, which is inductively seen to be a probability measure.

ductively by the equations $P_n^{(1)}(x;E)=P_n(x;E)$;
$P_n^{(s)}(x_0;E)=\int P_{n-1}^{(s-1)}(x;E)dP_n(x_0;x)$.

Corollary 2: For every set of absolute probabilities $Q_n(E)$ corresponding to $P_n(x;E)$ we have the equation $Q_{n+s}(E)=\int P_n^{(s)}(x;E)dQ_n(x)$.

We mention here three types of Markoff processes.

(i) Let X be the space of integers $1,2,\ldots,r$ and let B be the class of all subsets of X. Let $p=\{p_{ij,n}\}$ be a set of square matrices of order r satisfying (a) $\sum_j p_{ij,n}=1$ (b) $p_{ij,n}\geq 0$. We define $P_n(x;E)=\sum_{j\in E} p_{xj,n}$. We shall see in section II that there always exists at least one set of absolute probabilities in this case. It then follows from Theorem 1 that there always exists at least one Markoff process corresponding to any set of matrices satisfying (a) and (b). Such a process is called a finite process and, in the case where $p_{ij,n}$ is independent of n, has been studied by innumerable writers since Markoff.

(ii) Let X be any space and let m be any finite measure defined on a Borel field B of subsets of X. Let $p(x,y)$ be any function defined on $X\times Y$, the space of all pairs (x,y) where $x,y\in X$, and satisfying (a) $\int p(x,y)dy=1$ for every x (b) $p(x,y)\geq 0$. Define $P_n(x;E)=\int_E p(x,y)dy$. Under certain restrictions on the function $p(x,y)$ there will exist a function $p(x)$ satisfying the conditions $p(x)\geq 0$; $p(y)=\int p(x)p(x,y)dx$; $\int p(x)dx=1$. Then, writing $Q_n(E)=\int_E p(y)dy$, it is easily verified that $Q_n(E)$ is a set of absolute probabilities, and that the Markoff process thus determined is temporally homogeneous.

(iii) Let X be any space, B be any Borel field of subsets of X, and suppose that for each $x\in X$ a probability measure $P(x;E)$ is defined on B. We define $P_n(x;E)=P(x;E)$. Clearly this case includes case (ii). Under various hypotheses on the $P(x;E)$ results have been obtained in this case by Doob, Doeblin, Yosida and Kakutani, and other writers.

It is not known whether (iii) includes the general temporally homogeneous case, i.e. whether in any temporally homogeneous Markoff process conditional probabilities can be defined in such a manner that for every fixed x_n, $P(x_n;\{x_{n+1}\in E\})$ is a probability measure on B. If B satisfies a certain denumerability condition, however, Doob' has shown that the conditional probabilities can be so defined.

II. Matrix treatment of the finite case. In this and the following section we shall consider in detail processes of type (i). Let p_n be a set of square matrices of rank r as in section I. We shall use small letters a,p,l etc. to denote square matrices of rank r whose elements are a_{ij}, p_{ij}, l_{ij} etc. $a_n\to l$ means that for each i,j, $a_{ij,n}\to l_{ij}$ and we say that the sequence of matrices a_n converges to the matrix l. ab denotes the ordinary matrix product. We define $p_n^{(s)}$ inductively by the equations $p_n^{(1)}=p_n$; $p_n^{(s)}=p_n p_{n+1}^{(s-1)}$. Corollary 1 of Theorem 2 states that $p_n^{(s)}p_{n+s}^{(t)}=p_n^{(s+t)}$, which is easily verified directly in this case.

By the Cantor diagonal process, choose a sequence $n_\nu\to\infty$ such that $p_{n_\nu+a}^{(n_\nu+a)}$ converges for each value of a, say $p_{n_\nu}^{(n_\nu+a)}\to q^a$. Then $q^a p_a^{(b)}=q^{a+b}$.

A set of absolute probabilities may be represented by a set of 1 row, r column matrices $P^a=(P_1^a,\ldots,P_r^a)$, where a runs through all positive and negative integers and $P_j^{a+1}=\sum_i P_i^a p_{ij,a}$ (in matrix notation $P^{a+1}=P^a p_a$). The equation $q^a p_a=q^{a+1}$ states that, for any fixed subscript i, $q_i^a=(q_1^a,\ldots,q_r^a)$ is a set of absolute probabilities.

Choose a subsequence of n_ν for which q^{n_ν} converges, say $q^{n_\nu}\to l$. Then $lq^a=q^a$ and $l^2=l$. By a result of Doob", the last equation implies that the subscripts $1,\ldots,r$ may be divided into groups

---

' II. p. 96. It is unnecessary to exclude a set of measure zero in the statement of the theorem.

" IV.

$\bar{G}_1,\dots,\bar{G}_\rho,\bar{F}$ satisfying
(a) $l_{ij}=l_{ji}>0$ for $i,j\in\bar{G}_s$
(b) $l_{ij}=0$ for $i\in\bar{G}_s$, $j\notin\bar{G}_s$
(c) $l_{ij}=0$ for $i,j\in\bar{F}$.

Theorem 3: Each matrix $q^{(\omega)}_{ij}$ is divided into parts with identical rows by the groups of subscripts $\bar{G}_1,\dots,\bar{G}_\rho$; i.e. for $i,j\in G_s$, $q^{(\omega)}_{ik}=q^{(\omega)}_{jk}$. We write $q_{ik}=p_{sk}$ for $i\in\bar{G}_s$.

k sets of absolute probabilities $Q^{(\omega)}_1,\dots,Q^{(\omega)}_k$ are said to be linearly dependent if there exist $c_1,\dots,c_k$, not all zero, such that $\sum c_i Q^{(\omega)}_j=0$ for $j=1,\dots,r$ and all ω.

Theorem 4: $P^{(\omega)}_{ij}$, $i=1,\dots,\rho$ are linearly independent (i.e. not linearly dependent) sets of absolute probabilities, and every set of absolute probabilities is linearly dependent on these ρ sets.

III. Conditional probability treatment of the finite case.

Theorem 8: Let S be a cylinder set over $x_{n-1},x_{n-2},\dots$ on a Markoff process. Then, with probability 1, for $k\le s\le s+r<n$, $E(P(x_k;S),\dots,P(x_s;S);P(x_{s+1};S))=P(x_s;S)$.

The following theorem was proved by Doob.[1]

Theorem 6: Let $\dots,x_{-n},\dots,x_0,\dots,x_n,\dots$ be a sequence of functions such that, with probability 1, $E(x_m,\dots,x_n;x_k)=x_n$ for $m\le n\le k$. Then x_{-n} converges with probability 1 as $n\to\infty$.

We now consider the Markoff process Ω determined by the matrices p_n with the measure $P(E)$ determined by the set of absolute probabilities $P^{(\omega)}_j=1/\rho(P^{(\omega)}_{1j}+\dots+P^{(\omega)}_{\rho j})$. Since every set of absolute probabilities is linearly dependent on $P_1,\dots,P_\rho$, $P(E)$ has the property that if a set has P-measure zero, it has measure zero for every measure determined by any set of absolute probabilities.

An easy consequence of Theorems 5 and 6, applied to the process Ω, is that, for fixed j, $P(x_{-n};\{x_0=j\})$ converges with probability 1 as $n\to\infty$, say $P(x_{-n};\{x_0=j\})\to q^{(\omega)}_j$. By excluding a set of measure zero for each fixed j,m, we obtain a set Ω_0 such that $P(\Omega_0)=1$, $q^{(\omega)}_j$ exists for all j,m and all $\omega\in\Omega_0$.

Theorem 7: For every $\omega\in\Omega_0$ there exists a subscript i such that $q^{(\omega)}_j=q^{(\omega_i)}_j$.

We define $\omega_1\sim\omega_2$ if $q^{(\omega_1)}_j=q^{(\omega_2)}_j$. This is clearly an equivalence relation, and divides Ω_0 into at most ρ classes of equivalent points. Denote by $G_1,\dots,G_\rho$ those classes which have positive measure. Denote by $P_{G_i}(E)$ the measure determined by the set of absolute probabilities $q^{(\omega)}_j$ for $\omega\in G_i$.

We define subsets H_i of Ω_0 as follows:
$H_i=\{\omega\in\Omega_0$ such that $x_{-n}\in\bar{G}_i$ for an infinite number of $n\}$. Clearly for all $\omega\in H_i$, $q^{(\omega)}_j=P^{(\omega)}_{ij}$. It follows that $H_i H_j=0$ for $i\neq j$. The following theorem and corollary show the relation between the development in this and the preceding section.

Theorem 8: $P_{G_i}(H_i)=1$, where $P_{G_i}(E)$ is the measure determined by the set of absolute probabilities $P^{(\omega)}_{ij}$.

Corollary: $\rho=\rho_0$. By properly relabeling the sets G_i we have $G_i\supset H_i$, $P(G_i-H_i)=0$.

IV. The finite temporally homogeneous case. In this section we first state some results about general Markoff processes and then apply them, together with the results of the preceding section, to obtain, by a new method, the well-known results for finite temporally homogeneous Markoff processes. The following theorem states that a Markoff process is also a Markoff process going backwards, a fact which has been stated for finite processes by many writers.

Theorem 9: If Ω is a Markoff process, then for every cylinder set S over $\dots,x_{m-2},x_{m-1}$, $P(x_m,\dots,x_n;S)\,P(x_m;S)$ with probability 1.

Theorem 10: Let Ω be a Markoff process and let S be a cylin-

---

[1] II, p. 458.

der set over $\ldots, x_{n-1}, x_n$ for every n with characteristic function f. Then there exists a sequence of sets S_n with characteristic functions f_n such that S_n is a cylinder set over x_{-n}, $f_n \to f$ with probability 1 as $n \to \infty$.

Theorem 11: If Ω is a finite temporally homogeneous Markoff process and S is a cylinder set over $\ldots, x_{n-1}, x_n$ for every n, then S is, up to a set of measure zero, a cylinder set over x_n for each n, and there exists a positive integer j such that $T^j S = S$ up to a set of measure zero.

Let Ω be the Markoff process determined by a single matrix p with $P(E)$, $q_{ij}^{(n)}(\omega)$, $G_1, \ldots, G_\rho$ defined as in section III. It is easily verified that $q_{ij}^{(n)}(\omega) = q_{ij}^{(n)}(T\omega)$. This implies that $TG_i = G_j$ for some j, i.e. T merely induces a permutation among the G_i, and implies also that $P_{G_i}(E) = P_{TG_i}(TE)$. Then $P(E) = 1/\rho(P_{G_1}(E) + \ldots + P_{G_\rho}(E)) = 1/\rho(P_{TG_1}(TE) + \ldots + P_{TG_\rho}(TE)) = P(TE)$, and Ω is temporally homogeneous. It follows that for each i, G_i is, by Theorem 11, equal almost everywhere to a cylinder set over x_0, say $G_i = \{x_0 \in \overline{G}_i\}$ with probability 1. We may assume that $\overline{G}_i$ contains no subscripts of zero probability, i.e. that $P\{x_0 = j\} > 0$ for every $j \in \overline{G}_i$. Then $\overline{G}_1, \ldots, \overline{G}_\rho$ are disjunct, and $P\{x_0 \notin \overline{G}_1 + \ldots + \overline{G}_\rho\} = 0$. Write $\overline{F} = \{j \mid \overline{G}_1 + \ldots + \overline{G}_\rho\}$. By excluding a set of measure zero from each $\overline{G}_i$ we may assume that $x_n(\omega) \in \overline{G}_1 + \ldots + \overline{G}_\rho$ for all n and all $\omega \in G_i$. By Theorem 14 there exists a smallest positive integer s such that $T\overline{G}_1 = \overline{G}_1$ with probability 1. Then $G_1, TG_1, \ldots, T^{s-1}G_1$ are distinct sets G_i, and $I = G_1 + \ldots + T^{s-1}G_1$ is invariant' under T. In this manner we determine invariant sets $I_1, \ldots, I_\kappa$ such that each G_i is contained in exactly one set I. We shall for convenience relabel the G_i as follows: $I_m = G_{m1} + G_{m2} + \ldots + G_{ma_m}$ where $TG_{mi} = G_{mi}$ (except that $TG_{ma_m} = G_{m1}$), and shall use the corresponding notation for subscripts: $G_{mi} = \{x_0 \in \overline{G}_{mi}\}$, $TG_{mi} = \{x_0 \in T\overline{G}_{mi}\}$. The following theorem contains the principal results:

Theorem 12: (a) $p_{ij}^{(a_m n + d)} \to q_{ij} > 0$ as $n \to \infty$ for $i \in \overline{G}_{mr}$, $j \in T^d \overline{G}_{mr}$.
(b) $p_{ij}^{(a_m n + d)} = 0$ for $i \in \overline{G}_{mr}$, $j \notin T^d \overline{G}_{mr}$.
(c) $p_{ij}^{(n)} \to 0$ as $n \to \infty$ for $i, j \in \overline{F}$.

V. The equation $P(x_0; E) = \int P(x; E) dP(x_0; x)$. Let B be a Borel field of subsets of an abstract space X. Suppose that for each $x \in X$ a probability measure $P(x; E)$ is defined on B satisfying

(1) $$P(x_0; E) = \int P(x; E) dP(x_0; x).$$

Definition: A set N is null if for all $x \in X$, $P(x; N) = 0$.

Definition: A set J is invariant if there exists a null set N such that for every $x \in J - N$, $P(x; J) = 1$.

Theorem 13: The class I of invariant sets is a Borel field.

Two Borel fields will be called equivalent if for every set A in either there is a set B in the other such that $AB^c + A^cB$ is a null set. A Borel field B is called atomic if there exists a class of subsets A_t such that $A_t \in B$; $\sum A_t = X$; for every $S \in B$ and every t, either $A_t S = A_t$ or $A_t S = 0$. The sets A_t are called the atoms of B. A Borel field is called strictly separable if it is determined by a denumerable subcollection of its sets.

We shall at this point assume that I satisfies the condition (#). I contains an equivalent atomic subfield I_0 which contains a strictly separable subfield containing all atoms of I_0.

We shall later state conditions on B which imply (#). Under this condition, let A be the atoms of I_0.

Theorem 14: Under condition (#) there exist disjunct sets N, $B_t \in B$ such that (a) $X = N + \sum B_t$; $B_t \subset A_t$ for each t; $A_t - B_t$, N are null sets.

(b) $P(x; B_t) = 1$ for every $x \in B_t$.
(c) $P(x_1; E) = P(x_2; E)$ for every $x_1, x_2 \in B_t$.

VI. Markoff processes satisfying $P(x_0; E) = \int P(x; E) dP(x_0; x)$. In this section we shall consider two types of Markoff processes for which condition (#) is satisfied and to which, therefore, we may apply the results of the preceding section. Again we consider a

---

' i.e. TI = I with probability 1.

Borel field B of subsets of an abstract space X and suppose that for each $x \in X$ a probability measure $P(x;E)$ is defined on B satisfying (1). There always exist Markoff processes for which $P_n(x;E) = P(x;E)$; for (1) states that for any fixed x_0, $Q_n(E) = P(x_0;E)$ is a set of absolute probabilities.

We define a metric on B as follows: $d(A,B) = l.u.b.P(x; AB^{-}+A^{-}B)$. We shall say that B is uniformly separable if it contains a denumerable dense set with respect to the above metric.

Theorem 15: If B is uniformly separable, it contains an equivalent strictly separable subfield.

Now if B is uniformly separable, then the subfield I of invariant sets is also uniformly separable and hence, by Theorem 15, condition (#) is satisfied. We may therefore state

Theorem 16: If B is uniformly separable, we may write $X = N + \sum B_\tau$, where $P(x;E) = P(t;E)$ for all $x \in B_\tau$, $P(t;B_\tau) = 1$; and $P(x;N) = 0$.

Now suppose $\mathfrak{L}$ is a process of type (ii) for which $p(x,y)$ satisfies the relation

$$p(x,z) = \int p(x,y)p(y,z)dy. \tag{2}$$

Then, integrating both sides with respect to z over a set E and applying Theorem 2, we obtain (1). Suppose B is separable with respect to the measure $m(E)$, i.e. suppose B contains a denumerable dense subcollection with respect to the metric $d(A,B) = m(AB^{-}+A^{-}B)$. Then I is also separable with respect to m and contains a strictly separable subfield I_0 which is equivalent to I with respect to m. This implies that I_0 is equivalent to I with respect to the metric $d(A,B) = l.u.b._{x \in X} P(x;AB^{-}+A^{-}B)$ and hence that (#) is satisfied.

Theorem 17: If $\mathfrak{L}$ is a process of type (ii) for which $p(x,y)$ satisfies (2), and if B is separable with respect to m, then there exist sets $N, S, A_1, A_2, \ldots$ in B such that

(a) $X = N + S + A_1 + A_2 + \ldots$
(b) $p(x,y) = p_i(y)$ for $x \in A_i$.
(c) $p_i(y) = 0$ for $y \notin A_i + S$, $m(S) = 0$.
(d) $p_i(y) \neq 0$ for $y \in N$.

We remark that the set N need not be of m-measure zero. From (c), however, it is clearly a null set in the sense defined in section V.

Bibliography

I. Doeblin, W. Exposé de la theorie des chaînes simples constantes de Markoff à un nombre fini d'états. Revue Mathématique de l'Union Interbalkanique. 2 (1938) pp. 1-29

II. Doob, J. L. Regularity properties of certain families of chance variables. Transactions of the A. M. S. 47 (1940) pp. 455-486

III. Doob, J. L. Stochastic processes with an integral-valued parameter. Transactions of the A. M. S. 44 (1938) pp. 87-150

IV. Doob, J. L. Topics in Markoff chains. (unpublished)

V. Halmos, P. R. On the decomposition of measures. To appear in the Duke Mathematical Journal.

VI. Kolmogoroff, A. Grundbegriffe der Wahrscheinlichkeitsrechnung Ergebnisse der Mathematik, 2 no. 3 (1933)

VII. Kolmogoroff, A. Zur Theorie der Markoffschen Ketten, Math. Annalen 112 pp. 155-160

VIII. Levy, P. Theorie de l'Addition des Variables Aléatoires, Paris, 1937

IX. Saks, S. Theory of the Integral, New York, 1937

VITA

David Harold Blackwell was born on April 24, 1919, in Centralia, Illinois. He received his elementary and secondary school education in Centralia and entered the University of Illinois in September, 1935. He received the A.B. and A.M. degrees from the University of Illinois in 1938 and 1939 and was a fellow in mathematics there in 1939-40 and in 1940-41.